

SONIK: Efficient In-situ All Item Rank Generation using Bit Operations

Sourav Dutta

Max Planck Institute for Informatics,
Saarbrücken, Germany.
sdutta@mpi-inf.mpg.de

Abstract

Sorting, a classical combinatorial process, forms the bedrock of numerous algorithms with varied applications. A related problem involves efficiently finding the corresponding ranks of all the elements – catering to rank queries, data partitioning and allocation, etc. Although, the element ranks can be subsequently obtained by initially sorting the elements, such procedures involve $O(n \log n)$ computations and might not be suitable with large input sizes for hard real-time systems or for applications with data re-ordering constraints.

This paper proposes *SONIK*, a non-comparison linear time and space algorithm using bit operations inspired by *radix sort* for computing the ranks of all input integer elements, thereby providing implicit sorting. The element ranks are generated *in-situ*, i.e., directly at the corresponding element position without re-ordering or recourse to any other sorting mechanism.

1998 ACM Subject Classification F.2.2 Sorting and Searching

Keywords and phrases All Item Rank Computation, Integer Sorting, Non-comparison Sort, Bit Operations, Linear Time

1 Introduction and Related Work

Preliminaries: Given a collection of n integer elements, $\sigma = \{e_1, e_2, \dots, e_n\}$, *sorting* refers to the procedure of arranging the elements in accordance with a definitive ordering, e.g., numeric or lexicographic. Mathematically, sorting produces a permutation of the input items such that the elements in the final list are ordered as $e_i \leq e_j, \forall i \leq j$.

Efficient sorting algorithms provide a classical area of research in the domain of algorithms and beyond. The literature provides an enormous wealth of applications involving sorting, as a major operational step, such as comparison and searching, statistical analysis, and priority-based scheduling to name a few. Advanced algorithms such as knapsack problem [18, 7], minimum spanning tree [4, 9, 17], and network analysis also employ sorting as an intermediate stage. Several sorting algorithms such as bubblesort [16, 5], quicksort [14], mergesort [16], heapsort [21], etc., based on different strategies such as *divide-and-conquer*, *exchange*, *partitioning*, and *greedy approach* [5] have been proposed to cater to different applications and scenarios.

Sorting algorithms can be classified into two categories – *comparison sort* involving element-wise comparison or *counting based sort* using the count of items. Comparison sort such as mergesort, heapsort, etc., have been shown to be theoretically bounded by $O(n \log n)$ [5], where n represents the number of elements. On the other hand, counting-based sorts are bounded by $O(n)$ assuming certain characteristics of the input data such as uniform distribution (in bucket sort [5, 6]) or a small distribution range of the input element (in counting sort [8]). Further, implementational optimizations have also been studied for improving the practical efficiency of common sorting algorithms by using intelligent data structures, such as multi-key quicksort [3] and trie-based radix sort.



Sorting algorithms are also characterized by certain behavioral properties, such as *in-place* (involving the use of $O(1)$ extra storage space), *stable* (preserving the appearance order of duplicate elements in the sorted order) [5], and *adaptive* (taking advantage of pre-existing ordering in the input) [19]. The crucial nature of sorting techniques in various domains has also led to the development of several parallelized and disk-aware sorting procedures such as bitonic sort [2], odd-even sort [11], etc.

Motivation: An interesting and closely related problem involves the computation of the *ranks* of all the elements in an input list, where the rank of an item is defined as its position in the final sorted list. For example, given an input collection $\sigma = \{7, 2, 1, 5, 4\}$ the corresponding rank list for the elements is $\rho = \{5, 2, 1, 4, 3\}$, based on the non-decreasing key sorted order $\{1, 2, 4, 5, 7\}$ of σ . In fact, several applications involving scheduling, order-statistics, job priority, data partitioning, etc., [5, 16] use element ranks as the working metric. For example, consider the scenario of task scheduling from a job queue within a server pool with heterogeneous configurations, and the rank of a job to denote its priority. As such, the scheduling algorithm needs to compute the rank of all the jobs in the queue for the scheduler to appropriately allocate the job with the highest priority (i.e., highest rank) to the server with the greatest compute power and/or resource availability, and so on. Further, certain restrictions as to the re-ordering of items might apply in certain application scenarios. Hence, efficient approaches for computation of all item ranks to facilitate enhanced downstream processing provide an interesting area of study for such scenarios. Unfortunately, limited work exists in this regard, to the best of our knowledge.

Application: Consider an input list of n objects, where each object is represented by a d dimensional vector of integers corresponding to the values of its d dimensions. Applications such as *multi-criteria searching* require the input objects to be ordered such that most (or all) of the d -dimensional values are nearly (or completely) sorted. As such, the objects need to be independently sorted based on their i^{th} dimension to generate d such ordered lists. Finally, the d ordered lists are to be suitably combined into a final sorted list of the n objects such that most of the dimensions are nearly sorted. Observe that, an expensive join operation is required to be performed across the individual sorted lists (for each of the dimensions) as the different lists might have different ordering of the objects (based on the dimensional values). However, an efficient approach finding the ranks of all the objects for the individual dimensions in a pre-defined or their input appearance order would eliminate the need for the expensive merge procedure, thereby significantly improving the performance of the application. In this paper, we aim to tackle this *all-item ranking* problem by proposing an effective algorithm to compute a rank list containing the ranks of the input elements in the order of their appearance, i.e., in-situ.

Problem Statement: Formally, *given an input list of n elements, $\sigma = \{e_1, e_2, \dots, e_n\}$, the problem of all-item ranking entails the efficient computation of the rank list $\rho = \{r_1, r_2, \dots, r_n\}$, where r_i denotes the rank of element $e_i, i \in [1, n]$, with rank r_i referring to the position of element e_i in the sorted order of σ . We further aim to obtain the element ranks possibly without item re-ordering or use of sorting techniques which might be restricted in certain application scenarios (e.g., distributed job scheduling).*

State-of-the-art: Naïvely, we could initially sort the input list and thereafter obtain the ranks of individual elements by searching its position within the ordered list. Observe that, sorting of the input list can be performed in $O(n)$ best case by use of non-comparison based sorting approaches. However, such methods assume certain characteristics of the input elements (such as small range, uniform distribution, etc., as discussed earlier). Moreover, the computation of the rank list for the input elements (in the order of appearance in the

input list) involves subsequent searching of each of the elements in the sorted order for their positions (i.e., ranks). For example, in our previous example, for obtaining the rank of $e_1 = 7$ ($\in \sigma$) requires the searching of e_1 in the sorted list $\{1, 2, 4, 5, 7\}$. This involves additional $O(\log n)$ computations for each element, using binary search procedure. Hence, the total complexity for obtaining the rank list is bounded by $O(n \log n)$.

Certain scenarios might restrict the modification of the original order of elements, and hence re-ordering of elements for in-place sort may be infeasible (e.g., re-ordering of jobs and data allocated to servers might be forbidden or have significant network communication overheads). However, an additional $O(n)$ space to create a copy of the data for sorting is acceptable in most cases, as the generation of the rank list would also require $O(n)$ extra space.

The *quickselect* [15] algorithm working on the *divide-and-conquer* principle provides the state-of-the-art $O(n)$ approach to obtain the item with rank r from an unsorted list. However, finding the ranks of all elements degenerates its performance to quadratic complexity, i.e., $O(n^2)$, for finding the ranks of all items, i.e., $\forall r \in [1, n]$.

To alleviate such problems, the augmented *order-statistics tree* [5] was proposed to efficiently store and index the input elements in a tree-based structure by storing extra statistics at each node of the tree (e.g., number of descendants in the left subtree, etc.). Subsequent querying for the item having rank r can then be performed in $O(h)$, where h represents the height of the tree structure. The use of height-balanced tree data structures like *red-black tree* [10] or *AVL tree* [1] provides an upper bound of $O(\log n)$ for h and for finding the item with rank r . Thus, the construction and representation of input elements using an order-statistics tree and rank computation for all the elements (i.e., rank list) therein can be performed in $O(n \log n)$ time. However, such approaches suffer from involved tree rotational procedures for height balancing and also incur huge space requirements due to extra information and pointer storage within the nodes of the data structure.

Hence, we observe that state-of-the-art computation of the ranks of all input elements can be performed in $O(n \log n)$.

Contribution: In this paper, we propose the *SONIK* algorithm for the *all-item ranking* problem, i.e., finding the rank list for an input list of integer elements in their input appearance order, in $O(nk)$ time, where n is the number of input elements and k denotes the number of bits required to represent the elements. Considering, k to be constant (in the order of a few tens, since 32/64 bits are sufficient to encode huge numbers) we obtain a *linear* time algorithm, and show that the space requirement is also linear in n .

SONIK employs bit comparison operations (similar to the working principle of LSB radix-sort [13]) to provide an efficient and practical approach for real-time scenarios, and also an *anytime algorithm* [12] obtaining the partial ranks of elements based on the i least-significant bits after i iterations. We also show that no re-ordering of elements or dependence on other sorting approaches (as in radix-sort) is required by our approach leading to *in-situ* generation of the element ranks. Further, the generated ranks of the elements provide an inherent sorting of the input list in linear time.

Outline: The remainder of the paper is organized as follows: Section 2 describes the detailed working of the proposed *SONIK* algorithm. Section 3 discusses a few implementational optimizations along with avenues for generalization to handle different input scenarios and enforce sorting properties such as stability. Finally, Section 4 concludes the paper, followed by literary references in this problem domain.

2 SONIK Algorithm

In this section, we discuss the detailed working principle of our proposed *SONIK* algorithm. Without loss of generality, we assume the input list of elements, σ , to consist of non-negative integers without the presence of any duplicates, and that the item ranks are to be computed based on the non-decreasing sorted order of σ . Generalizations for handling other input characteristics and scenarios will be later presented in Section 3.

As a running example, consider $\sigma = \{7, 2, 1, 5, 4\}$ to be the input list and $\rho = \{r_1, \dots, r_5\}$ as the final element rank list to be computed, where r_i represents the rank of element $e_i \in \sigma$. Further, let k be the number of bits required to encode the elements in σ ; hence $k = 3$ for our above example.

The rank list, ρ is at first initialized to 1 and *SONIK* operates on the bit representations of the input elements, with processing proceeding from the least-significant bit (LSB) to the most-significant bit (MSB), i.e., from the rightmost to the leftmost bit. Thus, bit encoding of the input items, $\sigma_{bit} = \{111, 010, 001, 101, 100\}$ and initial $\rho = \{1, 1, 1, 1, 1\}$ forms the starting configuration for our algorithm. Starting from the LSB, *SONIK* iterates over the k bits of the input elements, and at each iteration performs two major computational phases – (1) *Rank Updation* and (2) *Rank Consolidation*.

At the onset of iteration i , the i^{th} least significant bits of all the elements in σ are extracted and based on the associated bit values, two lists, L_0^i and L_1^i , are created to store the current ranks of elements with their corresponding bits set to 0 and 1 respectively.

► **Observation 1.** Given only the least significant i bits of non-negative integers, elements with the i^{th} bit set (i.e., elements corresponding to the ranks in list L_1^i) have a higher numerical value compared to the elements associated with list L_0^i (i.e., having i^{th} bit value as 0), and hence should have larger ranks (based on non-decreasing sorted order of σ).

For our example, during the first iteration (i.e., considering only the least significant bits) elements $\{7, 1, 5\} \in \sigma$ have the rightmost bit set to 1 (from σ_{bit}) and hence their corresponding ranks are used to obtain list L_1^1 as $\{1, 1, 1\}$. Similarly, corresponding to the remaining elements, we obtain $L_0^1 = \{1, 1\}$ as shown in Figure 2. Based on the above observation, considering only the LSB, elements $\{7, 1, 5\}$ (of list L_1^1) should be assigned larger ranks as compared to elements $\{2, 4\}$ (in list L_0^1).

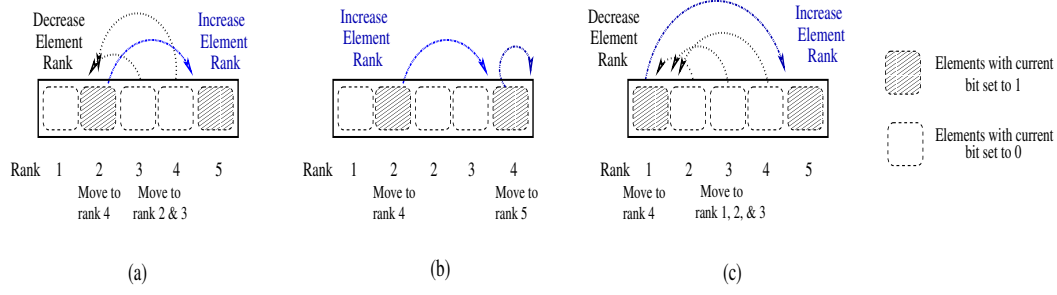
SONIK then performs the *rank updation* and *rank consolidation* phases separately for the elements in lists L_0^i and L_1^i , as discussed next.

2.1 L_0 List Processing

Initially, the ranks of the elements corresponding to the list L_0^i (obtained above) are appropriately processed and updated by the two stages of *SONIK* as follows:

Rank Updation Phase: Using the above computed lists, L_0^i and L_1^i , at each iteration the *rank updation* phase computes the following three operating variables:

- π_m^i :- denotes the *minimum* rank m among the elements having the i^{th} bit position set to 1, i.e., minimum value present in list L_1^i ;
- $\pi_{\leq m}^i$:- depicts the *maximum* rank *lesser than or equal to* π_m^i among the elements corresponding to list L_0^i , i.e., elements with i^{th} bit set to 0. The value of $\pi_{\leq m}^i$ is considered to be 0 if no such rank value is found; and,
- $\pi_{\geq m}^i$:- represents the *minimum* rank *greater than or equal to* π_m^i present in list L_0^i , i.e., among the ranks of elements with i^{th} bit set to 0. Alternatively, this value is set to $\pi_{\leq m}^i$ if no such rank is found.



■ **Figure 1** Working of the *Rank Updation* procedure in *SONIK* for various scenarios.

From Observation 1, based on the i^{th} bit, the elements corresponding to list L_1^i are numerically greater than the elements pertaining to list L_0^i , and hence their ranks need to be appropriately updated (increased). As such, the elements associated to list L_0^i having a rank greater than the minimum rank among the elements of L_1^i (π_m^i), are processed by *SONIK* during the rank updation stage. On the other hand, the ranks of elements of L_0^i with current ranks less than or equal to π_m^i remain unchanged during this step, as shown for the first element in Figures 1(a) and (b).

Observe that, in general the element(s) in L_0^i with rank $\pi_{\geq m}^i$ should now be re-assigned a rank of π_m^i , as it represents the currently numerically smallest element that was previously assigned a higher rank than π_m^i (in list L_1^i) during the iterations based on $i - 1$ LSB. Hence, in Figure 1(a), the item with rank 3 should now be assigned a rank of 2.

However, no such comparative deductions can be made for elements in L_0^i that have the same rank value of π_m^i (i.e., when $\pi_{\leq m}^i = \pi_m^i$). In these cases, the rank of such element(s) is kept unchanged during the current iteration and its final rank is computed by subsequent iteration steps. Hence, in unison, items in L_0^i with ranks less than or equal to π_m^i undergo no rank re-assignment as they are appropriately represented by their current ranks. In Figure 1(b), elements having rank 2 and 3 (depicted with unfilled boxes) represent such scenarios with unchanged current rankings.

Hence, for each element corresponding to list L_0^i having a rank greater than π_m^i , *SONIK* decrements the current rank by the *rank update* parameter value, δ_0 , defined as follows:

$$\delta_0 = \begin{cases} 0 & \text{if } \pi_m^i = \pi_{\leq m}^i \\ \pi_{\geq m}^i - \pi_m^i & \text{otherwise} \end{cases} \quad (1)$$

Thus, the updated element ranks are obtained by,

$$\rho_i = \begin{cases} \rho_i & \text{if } \rho_i \leq \pi_m^i \\ \rho_i - \delta_0 & \text{otherwise} \end{cases} \quad [\forall i, e_i \in L_0^i] \quad (2)$$

The pseudo-code for the *rank updation* phase is presented in Algorithm 1, while Figure 1 provides a pictorial description of the rank re-assignment strategy under different scenarios. The updated ranks of the elements corresponding to L_0^i are then provided as input to the *rank consolidation* phase, wherein the ranks are consolidated to represent contiguous positions.

Rank Consolidation Phase: Given the updated element ranks as obtained above, consolidation stage simply processed the ranks and to make them contiguous, so as to eliminate “holes” in the element rankings obtained. For example, an input rank list $\tau = \{1, 2, 4, 6\}$ is consolidated to $\tau' = \{1, 2, 3, 4\}$. We now make the following observation to enable the efficient performance of the *consolidation* phase for the *SONIK* algorithm.

Algorithm 1: *Rank Updation Procedure for L_0^i , $RUL_0()$*

Require: Input Element List (σ), Rank List (ρ), Element set S_0 corresponding to list L_0^i , and Operating variables (π_m^i , $\pi_{\leq m}^i$, and $\pi_{\geq m}^i$)

Ensure: Re-assignment of ranks to elements corresponding to list L_0^i .

```

1:  $A \leftarrow \pi_m^i$ ,  $B \leftarrow \pi_{\geq m}^i$ , and  $C \leftarrow \pi_{\leq m}^i$ 
2: if  $A = C$  then
3:    $\delta_0 \leftarrow 0$ 
4: else
5:    $\delta_0 \leftarrow (B - A)$ 
6: end if
7: for each element  $e_i \in \sigma$  and  $e_i \in S_0$  do
8:   if  $\rho_i > A$  then
9:      $\rho_i \leftarrow \rho_i - \delta_0$ 
10:  end if
11: end for
12: Return  $\rho$ 

```

► **Observation 2.** The updated rank list for the elements pertaining to list L_0^i is in non-decreasing sorted order.

Since the ranks of all the candidate items (in L_0^i) are reduced by a common factor, the *rank update* parameter, during consolidation, the original ordering among these elements are inherently preserved. Further, since the re-assigned ranks are greater than π_m^i , and element ranks less than π_m^i are not modified, the sorted property is preserved for the updated rank list (as the original rank list ρ was initially sorted).

We now show that with the use of an extra storage space and two array position pointers, the *rank consolidation* phase can be performed in linear time in the input rank list size. An additional array E of size $2*n$ (described later in Observation 3) is created, and each element $t_i \in \tau$ is mapped to the array position $E[t_i]$. The array position pointer, p_1 provides the next free slot, while the other position pointer p_2 refers to the next rank to be consolidated. The rank value of $E[p_2]$ is then accordingly set at position $E[p_1]$ based on the value at $E[p_1 - 1]$. This procedure is repeated until all the input ranks are consolidated as consecutive values.

For our example list τ , we create and initialize $E[8] = \{1, 2, -, 4, -, 6\}$ with the pointers initially set as $p_1 = p_2 = 1$. Iterating over the structure E , we obtain $p_1 = 3$ (the first free position of E) and $p_2 = 4$ (position of the next rank after p_1). The array E is then updated as $E[p_1] = E[p_1 - 1] + 1$ and $E[p_2] = -$. Hence, E is modified to $\{1, 2, 3, -, -, 6\}$. Finally, the first $|\tau|$ elements of $E = \{1, 2, 3, 4, -, -\}$ are returned as the consolidated rank list for τ . Algorithm 2 provides the pseudo-code for the rank consolidation procedure. Observe that, the sorted order of the input rank list enables the rank consolidation procedure to operate with a single pass of the array E , thereby running in linear time in the size of τ .

2.2 L_1 List Processing

After the processing of elements of list L_0^i , the ranks of the corresponding items are appropriately updated and consolidated to reflect their current ranks based on the i least significant bits. The ranks of the remaining elements corresponding to list L_1^i are now adjusted for assignment of higher ranks (as they are numerically greater) based on max , the maximum rank assigned to an element during the above rank consolidation procedure on list L_0^i . Observe that, currently the minimum ranked element in L_1^i (having a rank of π_m^i) should now

Algorithm 2: *Rank Consolidation Procedure, RCP()***Require:** Updated rank list τ of elements in list L_j^i from the rank updation procedure**Ensure:** Consolidated consecutive ranking for elements in τ

```

1:  $n \leftarrow |\tau|$ ,  $p_1 \leftarrow 1$ , and  $p_2 \leftarrow 1$ 
2: Create array  $E$  of size  $2 * n$ 
3: for element  $t_i$  in  $\tau$  do
4:    $E[t_i] \leftarrow t_i$ 
5: end for
6: while  $p_1 \leq n \wedge p_2 \leq n$  do
7:    $p_1 \leftarrow$  next free position in  $E$ 
8:    $p_2 \leftarrow$  next rank value after position  $p_1$ 
9:   if  $p_1 = 1$  then
10:     $E[p_1] = E[p_2]$ 
11:   else
12:     $E[p_1] = E[p_1 - 1] + 1$ 
13:   end if
14:   Remove entry  $E[p_2]$ 
15: end while
16:  $max \leftarrow E[n]$ 
17: Return first  $n$  elements of  $E$  as the consolidated rank list of  $\tau$  along with  $max$ 

```

Algorithm 3: *Rank Updation Procedure for L_1^i , $RUL_1()$* **Require:** Input Element List (σ), Rank List (ρ), Element set S_1 corresponding to list L_1^i , max returned by consolidation phase on list L_0^i , and Operating variable π_m^i **Ensure:** Re-assignment of ranks to elements corresponding to list L_1^i .

```

1:  $A \leftarrow \pi_m^i$ 
2:  $\delta_1 \leftarrow (max + 1 - \pi_m^i)$ 
3: for each element  $e_i \in \sigma$  and  $e_i \in S_1$  do
4:    $\rho_i \leftarrow \rho_i + \delta_1$ 
5: end for
6: Return  $\rho$ 

```

be assigned a rank of $max + 1$, and hence the current ranks of all elements pertaining to list L_1^i are incremented by the update parameter, $\delta_1 = max + 1 - \pi_m^i$. Hence, $\rho_i = \rho_i + \delta_1$ ($\forall i, e_i \in L_1^i$) forms the *rank updation* step for elements of list L_1^i as shown in Algorithm 3.

The obtained output ranks for list L_1^i from the above rank update procedure are then provided to the *rank consolidation* step for generating contiguous rankings, similar to the procedure described in Section 2.1.

We now reason about the $2 * n$ size bound of the extra storage array during the rank consolidation phase.

► **Observation 3.** The maximum rank obtained by an element during the rank updation stage is bounded by $2 * n$.

Proof. Consider the current minimum rank π_m^i of an element in L_1^i to be 1, i.e., numerically it denotes the minimum element based on its $i - 1$ least significant bits. Observe that, at the start of the i^{th} iteration, the previous rank updation and consolidation steps produce an ordered contiguous rank list ranging from 1 to n , the number of input elements in σ . Assume list L_1^i to also contain the element with the maximum current rank n . Considering all other

Algorithm 4: *SONIK Rank Generation Algorithm*

Require: Input element list σ
Ensure: Element Rank list ρ capturing the ranks of elements in σ

- 1: $n \leftarrow |\sigma|$
- 2: Create rank list ρ of size n and initialize all entries to 1
- 3: $k \leftarrow$ number of bits required to encode elements of σ
- 4: **for** $i := 1$ to k **do**
- 5: Extract the i^{th} bit, b_i^j of element e_j of σ , for $j \in [1, n]$
- 6: Create empty lists L_0^i and L_1^i
- 7: **if** $b_i^j = 0$ **then**
- 8: Add current rank of element e_j , ρ_j^i to list L_0^i
- 9: **else**
- 10: Add current rank of element e_j , ρ_j^i to list L_1^i
- 11: **end if**
- 12: Compute operating variable π_m^i , $\pi_{\leq m}^i$, and $\pi_{\geq m}^i$ from lists L_0^i and L_1^i
- 13: *RANK UPDATION:* Perform *RUL0()* on list L_0^i with σ , ρ^i , and operating variables
- 14: *RANK CONSOLIDATION:* Call procedure *RCP()* for list L_0^i with the above output
- 15: *RANK UPDATION:* Execute *RUL1()* on list L_1^i with σ , ρ^i , and operating variables
- 16: *RANK CONSOLIDATION:* Call procedure *RCP()* for list L_1^i with the above output
- 17: **end for**
- 18: Output ρ as the final rank list of elements of σ

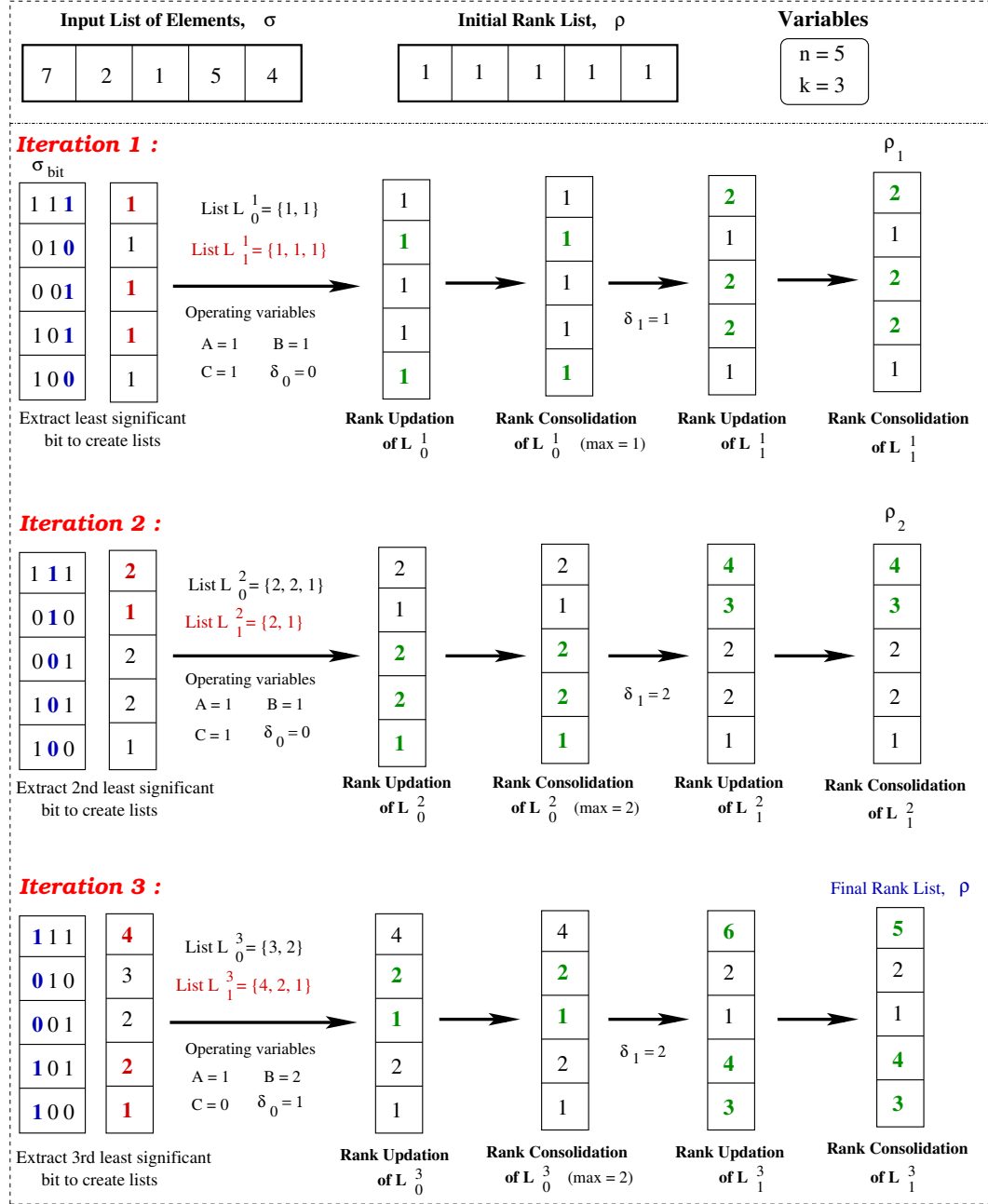
elements of σ to be in list L_0^i (as shown in Figure 1(c)) and the list to have been processed as described in Section 2.1, the maximum rank value, max returned by the consolidation procedure on L_0^i can be $n - 2$ (as list L_1^i is assumed to contain two elements).

In our current setting, the rank update parameter, $\delta_1 = max + 1 - \pi_m^i = n - 2 + 1 - 1 = n - 2$. The rank update procedure on list L_1^i might thus generate a maximum of $n + \delta_1 = 2 * n - 2$ as the updated rank for the maximum ranked element in L_1^i . Hence, the maximum rank assigned is upper bounded by $2 * n$, leading to the bound on the size of the extra storage array E for the rank consolidation phase.

Further, rank updation for list L_0^i decrements the current element ranks by the δ_0 parameter value. As the current maximum rank of any element can be n , the maximum rank obtained after the updation phase of *SONIK* remains bounded by $2 * n$. ◀

Overall Computation: The element rank list, ρ_i , obtained from the above computations (i.e., the rank updation and rank consolidation stages) by *SONIK* over the two lists L_0^i and L_1^i , depicts the intermediate ranking output at the end of the i^{th} iteration. The rank list, ρ_i is then fed as an input for processing during the next iteration step. This procedure is repeated until all the k bits of the input elements in σ have been processed, and the final rank list, ρ generated after the k^{th} iteration provides the output for our algorithm, containing the final ranks of all the input elements in the order of their appearance in σ . The pseudo-code for the complete working of the proposed *SONIK* algorithm is described in Algorithm 4. A working C++ code for *SONIK* can be obtained from <https://www.dropbox.com/s/zhn4q1hnh061mg6/SONIK.cpp?dl=0>.

Working Example: Continuing our previous example from Section 2, we now briefly walk through the operations of *SONIK* as depicted in Figure 2. For the input element list $\sigma = \{7, 2, 1, 5, 4\}$, during the first iteration, list L_1^1 contains the current ranks of elements $\{7, 1, 5\}$ (having LSB set to 1), while L_0^1 contains the ranks of the remaining items. Computing the



■ **Figure 2** Pictorial depiction of the working of *SONIK* on a running example.

operating variables, we obtain the update parameter $\delta_0 = 1$ for rank updation for list L_0^1 . Since the updated ranks returned are already contiguous, no change in the element ranks is perceived during the rank consolidation phase, which returns $\max = 1$ representing the maximum rank assigned during this step. The minimum rank of list L_1^1 should now be set to $\max + 1$, hence the update parameter $\delta_1 = 1$ is used for re-assigning the element ranks in list L_1^1 . Finally, after consolidation of L_1^1 , the obtained rank list at the end of the first iteration, ρ_1 forms the input for the next iterations, wherein a similar procedure is followed for the remaining bits of the input elements.

Finally observe that, during the last iteration, the rank updation of L_1^3 generates a rank list of $\{6, 4, 3\}$, for the elements $\{7, 5, 4\}$, based on the update parameter $\delta_1 = 2$. The rank consolidation phase is now seen to provide a contiguous rank list, removing ranking “holes” in L_1^1 (i.e., the absence of rank 5), to generate a consolidated rank list as $\{5, 4, 3\}$.

Discussion: *SONIK* thus provides an efficient linear time and space approach (discussed in Section 2.3) for computing the ranks (based on the non-decreasing sorted order) of all the input elements. Observe that our approach does not involve any item re-ordering (which might be restricted in distributed settings) or dependence on other sorting techniques.

Interestingly, *SONIK* depicts an *anytime algorithm*, wherein the rank computations can be terminated at the end of any iteration i , and the current rank list ρ_i provides the ranking of elements based on only the i least significant bits. This characteristic enables *SONIK* to effectively cater to hard real-time approximate ranking application scenarios.

Further observe that, at each iteration, the working of our algorithm involves the processing of only the i^{th} bits of all the input elements, without any other dependencies. As such, *on-demand* extraction of the appropriate bits incurs low network communication overhead (for items stored in distributed environments) at any particular time frame.

Finally, the output rank list for the input elements can be utilized to generate a sorted order of the input items. This can be performed in a single pass over the input and the rank lists (i.e., $O(n)$ time), by placing each element in the bucket with number corresponding to its computed rank, thus providing a linear time and space sorting algorithm.

2.3 Performance Analysis

In this section, we formally analyze the performance of the proposed *SONIK* algorithm for the *all item rank generation* problem. Assume the input element list σ to be composed of n integers, with each element represented by k bits.

From Algorithm 4, we observe that at iteration i , the i^{th} least significant bit of the input elements are extracted (ll. 5) and correspondingly the lists L_0^i and L_1^i are constructed, which can be performed in $O(n)$ time with a single pass over the input list, σ . Each element in σ contributes an entry in either of the two lists (ll. 7–11) and hence the combined size of the obtained lists is $|L_0^i \cup L_1^i| = |\sigma| = n$.

The two phases of *SONIK* – rank updation and rank consolidation – are invoked separately on the two lists (ll. 13–16), and hence each of the procedures operates on a total of n elements. The rank updation process iterates over each of the elements of the lists and appropriately updates the rank using the *updation parameters*, δ_0 and δ_1 . Thus, the time taken by the *rank updation* procedure is bounded by the total size of the two lists, i.e., $O(n)$.

The rank consolidation procedure, for each of the lists, creates an additional storage E of size $2 * n$ and consolidates (or removes “holes” from) the input element ranks by a single pass over E , with a time complexity of $O(n)$ as input rank is pre-sorted (as described earlier in Section 2.1). The computation of the operating variables can also be computed in linear time by two iterations over the lists L_0^i and L_1^i . Hence, the total time taken at each iteration i is bounded by $O(n)$ (compared to $O(n \log n)$ for state-of-the-art approaches), considering assignments and simple mathematical computations to be constant time operations.

Since, the total number of iterations is k (number of bits for input elements), the total running time complexity of *SONIK* thus becomes $O(nk)$. Considering k to a small constant (in the order of few tens) in most practical scenarios, the run-time of *SONIK* can be considered to be *linear* in the size of the input list (similar to the performance analysis of *radix-sort*) for which the element ranks are to be computed.

Also, the space complexity of *SONIK* is linear in the number of input elements, as the space required by the construction of lists and the extra storage during rank consolidation is bounded by $O(n)$. Hence, our proposed algorithm provides a time and space efficient real-time approach for computing the ranks of input elements without element re-ordering, use of sorting methods, or other input characteristic assumptions.

3 Generalizations and Optimizations

In this section, we discuss the generalizability of *SONIK* for handling different input element characteristics (such as negative values), and also possible avenues for optimizing the practical performance.

(1) Negative Elements: We initially assumed the input element list to be composed of non-negative numbers only. Modern computers store negative numbers using the *two's complement* binary representation technique [20]. In this approach, the most significant bit is set to 1 to denote a negative value. As such, during the last iteration step of *SONIK* (i.e., the k^{th} iteration), the negative elements will have their bits set and will mistakenly be considered to have a large numerical value and thus assigned larger ranks. To address such input scenarios, we initially perform an $O(n)$ pre-processing step, wherein the largest negative number present in the input (max_{neg}) is identified and all the input elements are then incremented by $|max_{neg}|$ value. This ensures that the input list provided to *SONIK* strictly contains non-negative elements.

(2) Duplicate Elements: In general, the working of *SONIK* inherently handles the presence of duplicate elements in the input list, and appropriately assigns the same rank to duplicate items. For example, the final element rank list $\rho = \{1, 2, 1\}$ will be generated corresponding to the input list $\sigma = \{2, 4, 2\}$. However, enforcing the *stability* criteria of sorting for obtaining the element ranks, the rank list for the above example should be $\{1, 2, 3\}$. To this extent, we use a variant of the counting sort procedure on the obtained element ranks as a post-processing step. The counting sort procedure operates on the generated rank list to provide a re-ordering of the ranks compliant with the stability property. Observe that, this approach requires $O(n)$ time and also $O(n)$ space as the final element ranks obtained range from 1 to n .

(3) Optimizing k : *SONIK* utilized the weighted bit representation of elements to compare the numerical values of the input elements and generate the final rankings. As such, a mapping of the original input items to a different set of elements preserving the sorted ordering among the initial elements, and the rank computation thereon also produces a correct ranking list. We use this intuition for possible reduction in the number of iteration steps, thereby improving the practical run-time performance of our algorithm. The input elements are pre-processed to identify the minimum item, min , and all the input numbers are subsequently decremented by a value of min . Observe that, this translation of the input items preserves the sorted order among the elements, and it is on these “new” items that the element ranks are computed. This process might help in possibly reducing k , the number of bits required to represent the input elements (as the maximum value among the input elements is reduced), thereby enhancing the practical performance of *SONIK*.

(4) Sorting Order: In scenarios where the required element rankings are to be based on the non-increasing sorted order (previously we considered non-increasing order), the operational stages and working of *SONIK* remain nearly unchanged, thus providing a robust framework. In such cases, only the generation of the lists L_0^i and L_1^i are reversed, that is, at iteration i , elements with the i^{th} bit set to 1 are considered for constructing the L_0^i list, while the other

elements are used for list L_1^i .

4 Conclusions

This paper presented *SONIK*, a *linear time and space* algorithm for computing the rankings (based on sorted order) for all elements of an input list. Our algorithm uses bit operations to compare the numerical values of elements for assigning appropriate ranks, based on two operational steps – rank updation and rank consolidation. The ranks are efficiently generated in the order of element appearance in the input list, and hence eliminates the need for item re-ordering or intermediate sorting as compared to state-of-the-art approaches. *SONIK* is also shown to demonstrate robust characteristics like *on-demand bit extraction* and *anytime algorithm*, also providing a linear-time sorting technique. Extension of our method to handle strings and other input data types provides an interesting area of future studies.

References

- 1 G. G. Adelson-Velsky and E. Landis. An algorithm for the organization of information. *USSR Academy of Sciences*, 146:263–266, 1962.
- 2 K. E. Batcher. Sorting Networks and their Applications. *AFIPS Spring Joint Computing Conference*, 32:307–314, 1968.
- 3 P. E. Black. Multikey Quicksort, 2005. Dictionary of Algorithms and Data Structures.
- 4 O. Boruvka. O Jistém Problému Minimálním (About a Certain Minimal Problem) (in Czech, German summary). *Práce Mor. Přírodoved. Spol. v Brně III*, 3, 1926.
- 5 T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press and McGraw-Hill, second edition, 2001.
- 6 E. Corwin and A. Logar. Sorting in linear time — variations on the bucket sort. *Journal of Computing Sciences in Colleges*, 20(1):197–202, 2004.
- 7 T. Dantzig. *Number: The Language of Science: A Critical Survey Written for the Cultured Non-Mathematician*. Macmillan, 1930.
- 8 J. Edmonds. Counting Sort (a Stable Sort), How to Think about Algorithms. *Cambridge University Press*, pages 72–75, 2008.
- 9 M. L. Fredman and R. E. Tarjan. Fibonacci heaps and their use in improved network optimization algorithms. *Journal of the ACM*, 34(3):596–615, 1987.
- 10 L. J. Guibas and R. Sedgwick. A Dichromatic Framework for Balanced Trees. In *Annual Symposium on Foundations of Computer Science*, pages 8–21, 1978.
- 11 N. Haberman. Parallel Neighbor Sort (or the Glory of the Induction Principle). Technical report, CMU Computer Science, 1972.
- 12 J. A. Hendler. *Artificial Intelligence Planning Systems*. Elsevier, 1992.
- 13 H. Herman. Art of Compiling Statistics, 1889. Patent US395781(A).
- 14 C. A. R. Hoare. Algorithm 64: Quicksort. *Communications of ACM*, 4(7):321–322, 1961.
- 15 C. A. R. Hoare. Algorithm 65: Find. *Communications of ACM*, 4(7):321–322, 1961.
- 16 D. Knuth. *The Art of Computer Programming*, volume Third: Sorting and Searching. Addison-Wesley, second edition, 1998.
- 17 J. B. Kruskal. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *American Mathematical Society*, 7(1):48–50, 1956.
- 18 G. B. Mathews. On the partition of numbers. *London Mathematical Society*, 28:486–490, 1897.
- 19 O. Petersson and A. Moffat. A framework for adaptive sorting. In *Lecture Notes in Computer Science*, volume 621, pages 422–433. Springer, 1992.

- 20 J. von Neumann. First Draft of a Report on the EDVAC. *IEEE Annals of the History of Computing*, 15(4):27–75, 1993.
- 21 J. W. Williams. Algorithm 232 - Heapsort. *Communications of ACM*, 7(6):347–348, 1964.